Week-6 ML Assignment

Ansh Harjai (ah7163)

Part – A

1. Prior and Likelihood Analysis (0.5 points)

Patients: 20

Syndrome X = Positive: 9 patients

Syndrome X = Negative: 11 patients

P(Syndrome X = Positive) = 9 / 20 = 0.45

P(Syndrome X = Negative) = 11 / 20 = 0.55

Likelihood of symptoms where patient is positive:

P(Fever = Yes | Positive) = 8 / 9 = 0.889

P(Joint Pain = Yes | Positive) = 6 / 9 = 0.6667

P(Rash = Yes | Positive) = 7 / 9 = 0.778

P(Fatigue = Yes | Positive) = 6 / 9 = 0.667

Likelihood symptoms where patient is negative:

P(Fever = Yes | Negative) = 2 / 11 = 0.18

P(Joint Pain = Yes | Negative) = 4 / 11 = 0.36

P(Rash = Yes | Negative) = 4 / 11 = 0.36

P(Fatigue = Yes | Negative) = 6 / 11 = 0.54

Applying Laplace smoothing with alpha = 1:

Smoothed likelihoods:

For syndrome = Positive

P(Fever = Yes | Positive) = 9 / 11 = 0.812

P(Joint Pain = Yes | Positive) = 7 / 11 = 0.636

P(Rash = Yes | Positive) = 8 / 11 = 0.727

P(Fatigue = Yes | Positive) = 7 / 11 = 0.636

For Syndrome = Negative

P(Fever = Yes | Negative) = 3 / 13 = 0.23

P(Joint Pain = Yes | Negative) = 5 / 13 = 0.38

P(Rash = Yes | Negative) = 5 / 13 = 0.38

P(Fatigue = Yes | Negative) = 7 / 13 = 0.538

Why Laplace smoothing is necessary:

Laplace Smoothing is necessary because it ensures that no probability becomes zero just because a symptom did not appear in one of the classes.

In medical diagnosis, small datasets can cause unseen symptom combinations to give zero probabilities. Laplace smoothing prevents this by adding a small correction term.

2. **MAP vs MLE Comparison (0.5 points)**

Maximum A Posteriori (MAP)

MAP combines data with prior medical knowledge.

If doctors know that only 5% of the population actually has Syndrome X, we use that as a prior:

P(Syndrome X = Positive) = 0.05

P(Syndrome X = Negative) = 0.95

Maximum Likelihood Estimation (MLE)

MLE relies only on the data. It chooses parameters that make the observed data most likely.

In our dataset, 9 out of 20 patients have Syndrome X, so

P(Syndrome X = Positive) = 9/20 = 0.45

P(Syndrom X = Negative) = 11/20 = 0.55

This assumes the sample perfectly represents the population.

Effect of the Prior

The MAP estimate adjusts predictions by giving more weight to this prior.

Even if symptoms suggest Syndrome X, the model will be cautious because the disease is rare.

PART – B

3. **Classification with Naive Bayes (1 point) A new patient arrives with the following symptoms:**

Given new patient:

Fever = Yes, Joint Pain = No, Rash = Yes, Fatigue = No

Use Laplace-smoothed likelihoods (α = 1) from Part A

Priors (from Part A):

P(Pos) = 0.45

P(Neg) = 0.55

Step 1 — Likelihood of the symptom vector under each class

Multiply the conditional probabilities (Naive Bayes assumption: symptoms independent given class).

Likelihood (symptoms | Positive)

= P(Fever=Yes|Pos) × P(Joint=No|Pos) × P(Rash=Yes|Pos) × P(Fatigue=No|Pos)

= 0.818 × 0.363 × 0.727 × 0.363

≈ 0.07868

Likelihood (symptoms | Negative) =

= 0.230 × 0.615 × 0.384 × 0.461

≈ 0.02521

Step 2 — Multiply by priors (compute unnormalized posteriors)

Unnormalized P(Pos | data) = P(Pos) × Likelihood(symptoms | Pos)

= 0.45 × 0.078

= 0.0354

Unnormalized P(Neg | data) = P(Neg) × Likelihood(symptoms | Neg)

= 0.55 × 0.025

= 0.0138

Step 3 — Normalize to get posteriors

Total = 0.035 + 0.0138 = 0.0492

P(Syndrome X = Positive | data) = 0.035 / 0.0492 **≈ 0.7186**

P(Syndrome X = Negative | data) = 0.0138 / 0.0492 **≈ 0.2814**

Classification decision

Predict Syndrome X = Positive.

Why: Posterior for Positive ≈ 71.9%, which is substantially higher than Negative ≈ 28.1%. Under Naive Bayes (with the Part A priors and Laplace-smoothed likelihoods) the evidence from the symptoms yields a clear positive prediction.

4. **Independence Assumption Analysis (1 point)**

Naive Bayes assumes that symptoms are independent given the diagnosis, meaning that once we know if a patient has Syndrome X, one symptom shouldn't affect another.

In reality, this isn't fully true. The symptoms in medicine often occur together because they share common causes.

Examples of dependent symptoms:

- Fever and Fatigue: Fever often causes tiredness, so they're linked even if we know the diagnosis.
- Rash and Joint Pain (or Fever): These can appear together in inflammatory or autoimmune conditions, so they're not truly independent.

Why Naive Bayes still works well:

- Each symptom still provides useful information.
- Small correlations don't always harm predictions.
- The model is simple, stable, and performs well even with limited data.